Constraints into Preferences: Gender, Status, and Emerging Career Aspirations

Shelley J. Correll
Cornell University

This study presents an experimental evaluation of a model that describes the constraining effect of cultural beliefs about gender on the emerging career-relevant aspirations of men and women. The model specifies the conditions under which gender status beliefs evoke a gender-differentiated double standard for attributing performance to ability, which differentially biases the way men and women assess their own competence at tasks that are career relevant, controlling for actual ability. The model implies that, if men and women make different assessments of their own competence at career-relevant tasks, they will also form different aspirations for career paths and activities believed to require competence at these tasks. Data from the experiment support this model. In one condition, male and female undergraduate participants completed an experimental task after being exposed to a belief that men are better at this task. In this condition, male participants assessed their task ability higher than female participants did even though all were given the same scores. Males in this condition also had higher aspirations for career-relevant activities described as requiring competence at the task. No gender differences were found in either assessments or aspirations in a second condition where participants were instead exposed to a belief that men and women have equal task ability. To illustrate the utility of the model in a "real world" (i.e., nonlaboratory) setting, results are compared to a previous survey study that showed men make higher assessments of their own mathematical ability than women, which contributes to their higher rates of persistence on paths to careers in science, math, and engineering.

How do gender differences in career choices emerge? Understanding the gendered nature of the career choice process is important since, to the extent that men and women make different career-relevant choices throughout their lives, the labor force will continue to be segregated by gender. Gender segregation in paid work is stubbornly resilient, persisting despite other structural changes in society, changes such as the vast movement of women into paid work in recent decades (Jacobs 1989, 1995a; Jacobsen 1994; Reskin 1993) and the transformation of work content due to technological changes and the increase in service sector jobs (England 1981; Game and Pringle 1983; Reskin and Roos 1990; Tienda and Ortiz 1987). The distribution of men and women into different kinds of occupations, firms, and establishments is consequential, explaining the majority of the gender gap in wages (Peterson and Morgan 1995).

Many explanations of this continued segregation have examined the impact of "demand-side" processes, a phrase referring to processes that lead to a greater demand for men when filling more desirable jobs (Anker 1997; England 1992; Nelson and Bridges 1999; Reskin and Roos 1990). This paper focuses

Direct all correspondence to Shelley J. Correll, Department of Sociology, Cornell University, 323 Uris Hall, Ithaca, NY 14853, sjc62@cornell.edu.
This project was made possible with support from the National Science Foundation (award no. SES-0000201) and the Institute for Research on Women and Gender, Stanford University. I wish to thank Cecilia Ridgeway and Myra Marx Ferree for their helpful suggestions and guidance.

instead on the "supply side" of the issue by addressing how men and women develop preferences or aspirations for different kinds of work.[1] Most scholars of gender inequality have been reluctant to develop supply-side explanations because these explanations often "blame the victim" (Browne and England 1997). However, by developing models that are truly sociological (i.e., that explicate how macro-level variables constrain individual action) it should be possible to understand how gender differences in career choices emerge without simultaneously suggesting that women voluntarily choose less advantageous positions in the labor market.

This paper, which is part of a larger project to develop a theory about gender and the career choice process, presents an experimental evaluation of one model that describes the constraining effects of cultural beliefs about gender on the emerging career-relevant preferences or aspirations of men and women. The main hypothesis is that cultural beliefs about gender that accord men higher status in society than women (i.e., status beliefs) can evoke gender-differentiated standards for attributing performance to ability, which differentially biases the assessments men and women make of their own competence at career-relevant tasks. This paper uses status characteristics theory and the empirical literature on stereotype threat to explain how and when this biasing effect is likely to occur. The implication is that, if individuals act on gender-differentiated evaluations of their own competence when forming aspirations for activities that lead to different careers, then status beliefs about gender will also differentially impact the career-relevant choices that men and women make. In the aggregate, if men and women systematically make different career-relevant choices, the gender-segregated labor force is necessarily reproduced. Before developing the model further, I briefly review supply-side explanations of gender segregation in paid work.

[1] In taking a supply-side approach, I am not suggesting that demand-side approaches are not important. To the contrary, this has been and continues to be a very important line of work. My argument is that demand-side process cannot *fully* account for gender segregation in paid labor (see England 1992 and Reskin and Roos 1990 for a review).

SUPPLY-SIDE EXPLANATIONS OF GENDER SEGREGATION

Early on the path to many careers, men and women—indeed, even boys and girls—begin to differentially commit themselves to activities that are career relevant. As early as high school, and even more strikingly by college, young men and women elect to take different kinds of courses and choose different college majors, which produces gender differences in the kinds jobs that are later seen as plausible options for students (AAUW 1992; Jacobs 1995b; National Science Board 1993; National Science Foundation 1994). Given this early gender divergence, it is probably not surprising that those who study labor market matching processes (i.e., the processes by which prospective employees become matched with employers) find that the supply networks from which employers recruit are highly segregated by gender (Granovetter and Tilly 1988). The gender segregation of job supply networks means that, even if all gender discrimination at the point of hire and subsequent promotion were removed, considerable gender segregation would still remain in paid work due to the different and seemingly voluntary career choices men and women make.

Human Capital Explanations

Scholars studying labor market matching processes tend to downplay the issue of gender differences in job supply networks. They assume that men and women have different tastes, preferences, or ways of maximizing utility, which leads to differences in men's and women's choices in careers and/or jobs. For example, human capital theorists have argued that women choose jobs with flatter rates of wage growth, because these jobs, which are primarily in female-dominated occupations, have smaller wage penalties for sustained periods of absence from the paid labor force and have higher starting wages (Polachek 1976, 1981; Zellner 1975). According to these theories, women know they will likely need to take an extended absence for child birth and/or care, so they choose jobs with the above characteristics to maximize their lifetime earnings. However, England and colleagues (1984; 1988) demonstrate that, contrary to the predictions of human capital theory, women employed

in male-dominated occupations actually have higher lifetime earnings.

When human capital theorists are confronted with evidence that men and women with equivalent human capital are found in jobs with different wages or different lifetime earning potential, they often expand their model of individual choice (Glass 1990). The most common expansion of this model is that women choose jobs that maximize their ability to coordinate family and paid work responsibilities (Marini and Brinton 1984; Polachek 1976). However, Glass (1990) shows that male-dominated jobs—compared with female-dominated jobs—are actually associated with more flexibility and autonomy, thus allowing a person, for example, to more easily leave work to tend to a sick child. In sum, women maximize neither earnings nor their ability to coordinate family and paid work duties by working in female-dominated occupations, leaving the question of why women and men choose different kinds of careers unanswered by the human capital perspective.

Cultural Constraints on Choice

What is needed is a supply-side approach that recognizes that the culture in which individuals are embedded constrains or limits what these individuals deem possible or appropriate, thereby shaping the preferences and aspirations that individuals develop for activities leading to various careers, often starting early in the life course. Pierre Bourdieu (1984 [1979:175]) articulated a most compelling explanation of this type of approach when he described how social class frames or constrains preferences and choices. Bourdieu noted that the "habitus," which includes an internalization of the power and status relationships between groups of people in a society, "continuously transforms necessities into strategies, constraints into preferences and, without any mechanical determination, it generates a set of 'choices' . . . It is a virtue made of necessity which continuously transforms necessity into virtue by inducing 'choices' which correspond to the condition of which it is the product."

This description is engaging, but it fails to specify *how* choices are induced. One goal of the current project is to develop a model that can account for this type of constraint on choice and that can be evaluated empirically. While there are undoubtedly many reasons why individuals develop preferences for one career or another, my model assumes that, as a minimum, individuals must believe they have the skills necessary for a given career in order to develop preferences for that career. I refer to a person's understanding of his or her own competence as a "self-assessment." The model explains how cultural beliefs about gender bias the formation of self-assessments of their competence at career-relevant tasks. I use "career-relevant" to refer to tasks, activities, decisions, and aspirations that, when performed, enacted, or held, impact the trajectory or path of an individual's job or career history. For example, going to graduate school is a career-relevant activity. Before developing the model further, I first describe a companion study in the larger project that shows the benefit of focusing on self-assessments of task competence when trying to understand how gender differences in career preferences or aspirations emerge.

Using a national probability sample of high school and college students, I measured the extent to which cultural beliefs about gender and mathematics contribute to the gender gap in careers in science, math, and engineering (Correll 2001). Research has shown that students, parents, and teachers perceive mathematic skills to be associated with masculinity and verbal skills are not (c.f., Hyde et al. 1990). Assuming that students in my sample were aware of these beliefs, I hypothesized that cultural beliefs about gender and mathematics lead men to make higher assessments of their own mathematical competence than women do. Controlling for grades and test scores in mathematics, I found that male high school students indeed rated their own mathematical ability (but not verbal ability) higher than female students did. Also, self-assessments of task competence impacted early career-relevant decisions: controlling for actual ability, the higher students assessed their own mathematical ability, the greater their odds of enrolling in a high school calculus course and choosing a college major in science, math, or engineering. Most importantly, when mathematical self-assessment levels were controlled, the previous higher enrollment of male students in a calculus course disappeared and the gender gap in college major choice was reduced.

Similarly, the psychologist Eccles and her colleagues (Eccles 1994; Eccles, Barber and Jozefowicz 1999) analyzed longitudinal survey

data from a sample of students in Michigan to evaluate part of their model of achievement-related choices. Their analyses show that the "cultural milieu" (which includes gender role stereotypes) influences parents' expectations of their children's abilities, which ultimately influences children's self-perceptions and their decisions regarding school course selection.

The results of the companion study (Correll 2001), and less directly the results of Eccles (1994; Eccles et al. 1999), provide evidence that is consistent with the main causal hypothesis that cultural beliefs about gender differentially bias men and women's self-assessments of task competence.[2] The companion study also uses "real world" data to show the utility of the model: self-assessments of task competence impact career-relevant decisions. However, these results can only provide limited support for the more general theoretical model for several reasons.

First, although the results of the companion study were obtained using a probability sample, they are still specific to a very limited set of tasks and professions, thereby limiting the generality of the model. Second, while the results illustrate the *impact* of self-assessments, they do not definitively support the hypothesis that cultural beliefs about gender bias their formation. This is because we must assume that the students in the sample were aware of the cultural beliefs about gender, mathematical abilities, and verbal abilities, and this awareness caused the observed gender differences in self-assessments of competence. Since we could not isolate and manipulate students' exposure to gender beliefs associated with these abilities, we could not rule out competing explanations for the findings. For example, the companion study could not address the possibility that mathematical self-assessments tap an additional component of "real" mathematical ability not captured by math grades and test scores that served as controls of mathematical ability. If this is correct, the higher self-assessments male students make of their mathematical ability might not be the result of cultural beliefs about gender and mathematics that exist in society, but might instead emerge because men "really are better" at mathematics. To the extent that measures of ability are imperfect, statistical controls of ability cannot rule out explanations of unmeasured actual ability.

The following social psychological experiment was designed to overcome these limitations and provide evidence that allows for a more definitive evaluation of how gender status beliefs bias self-assessments of task competence, as described below. However, each type of data (i.e., the data from the experiment and data from the probability sample) has its own strengths. Therefore, the theoretical argument is best evaluated by simultaneously considering the results of these two companion studies. Before describing the experiment, I first develop the theoretical model.

STATUS BELIEFS, SELF-ASSESSMENTS, AND EMERGING ASPIRATIONS

I rely on status characteristics theory and the empirical literature on "stereotype threat" to develop a model that describes the constraining effect of cultural beliefs about gender on self-assessments of task competence and emerging career-relevant aspirations. Status characteristics theory has developed empirically supported propositions linking cultural beliefs about social categories (such as gender and race) with inequalities in participation and evaluations in collectively oriented task groups (i.e., groups who work together to complete a joint task, such as work groups or athletic teams) (Berger et al. 1977). The literature on stereotype threat suggests that the impact of these beliefs on the reproduction of categorical inequality might occur in a wider range of settings than collectively oriented task groups, including situations where individuals evaluate or assess their own task competence in non-group settings.

Status Characteristics Theory

An attribute that differentiates people is a *status characteristic* if there are widely held beliefs

[2] Both Eccles (1994) and Correll (2001) suggest that cultural beliefs about gender impact individuals' self-perceptions and career or educational decisions, though their causal mechanisms differ. For a more detailed comparison, see Correll (2001). For a more general description of how psychological sex role socialization explanations differ from situational approaches, see Wagner and Berger (1997). The model and method presented in the current study have advantages over both Eccles (1994) and Correll (2001).

in the culture attaching greater social value and competence with one category of the attribute (men, computer expert) than another (women, computer novice) (Berger et al. 1977). Although status characteristics, as defined by the theory, have beliefs about the relative competence of social categories attached to them, whether or not these beliefs generalize from the *category* to a *member* of the category in any particular instance depends on features of the setting. The theory is composed of propositions about when this generalization will occur and how it leads to inequalities in participation and evaluations (Berger et al. 1977; Lovaglia et al. 1998; Troyer and Younts 1997; Webster and Foschi 1988).

Status characteristics theory has primarily restricted its scope to collectively oriented task groups because in these groups, the shared pressure to successfully complete the group's task causes group members to unconsciously anticipate the *relative* quality of each member's future task performances. When a status characteristic is salient for individuals in a setting, as defined below, its cultural association with greater or lesser worthiness and competence shapes the implicit and relative performance expectations members form for one another. Those possessing the more valued state of a salient characteristic are expected to offer more valuable task contributions in the group than those with the less valued state. Therefore, in a self-fulfilling manner, higher status individuals are given more opportunities to participate and, when they do participate in the group, their contributions are evaluated more positively. Experiments confirm that a wide variety of status characteristics systematically organize influence, participation, and the appearance of competence in this manner (see Webster and Foschi 1988).

The strength of status characteristics theory is that in addition to demonstrating that status characteristics influence behaviors and evaluations, it also specifies the circumstances under which status characteristics have their effect and the relative strength of their impact under differing conditions. For example, a status characteristic is only predicted to impact behavior or evaluation when it is salient in the setting. A status characteristic is salient when it differentiates those in a group (e.g., gender is salient in a mixed-sex group) or when it is believed to be relevant to the group's focal task (e.g., computer expert/novice might be relevant to a group whose goal is to design a web page) (Wagner and Berger 1993). The theory further argues that any salient status characteristic will shape expectations of those in the setting unless it is specifically dissociated from the task at hand. That is, the burden of proof lies with establishing that the characteristic is *not* relevant. However, the more directly relevant to the task the characteristic is perceived to be, the greater the strength of the impact of the status characteristic on behaviors and evaluations. For example, given the content of the beliefs about gender in our society, we might expect gender to seem more relevant and therefore have a larger impact on behavior and evaluations in a group repairing a car than in a group planning a fund-raising event. I will return to this point about the relevance of the characteristic to the focal task since it will be important for establishing how and when cultural beliefs about gender influence the formation of self-assessments of task competence when individuals are not in group settings.

Gender as a Status Characteristic

Gender is commonly described as a *diffuse* status characteristic, meaning that widely shared cultural beliefs about gender include expectations that men are diffusely more competent or capable at most things, as well as specific assumptions that men are better at some particular tasks (e.g., mechanical tasks) while women are better at others (e.g., nurturing tasks) (Conway, Pizzamiglio, and Mount, 1996; Fiske et al. 2002; Wagner and Berger 1997; Williams and Best 1990). Beliefs about gender and competence have changed over time; however, empirical studies continue to find that men are thought to be generally more capable (Williams and Best 1990:334) and competent (Fiske et al. 2002:892) than women. For example, Fiske et al. (2002:892) surveyed nine diverse samples, from different regions of the United States, and found that members of these samples, regardless of age, consistently rated the category "men" higher than the category "women" on a multidimensional scale of competence.

Since my argument draws on the psychological literature on stereotypes, it is worth comparing gender stereotypes to gender status beliefs. Gender stereotypes are often conceptualized as a broad set of beliefs about the kinds of traits, attributes, or behaviors that can be (or

should be) expected of a person of a given sex category (Deaux and Kite 1987). By contrast, gender status beliefs are a specific component of gender stereotypes: they are beliefs that men are more socially valued and diffusely more competent than women at things that "count" (Wagner and Berger 1997; Ridgeway and Correll 2000). Gender stereotypes include status beliefs, but also include other types of beliefs, such as beliefs about appropriate role behaviors, sexuality, and violence.

Both sociologists and psychologists often note that gender beliefs reflect a cultural system, representing what we think "most people" believe or accept as true about the categories of "men" and "women" (Ridgeway 1997; Deaux and Kite 1987). As such, status beliefs, and stereotypes more generally, operate as schemas for interpreting and making sense of the social world (Fiske 1998; Ridgeway 1997). Viewing status beliefs as cultural schemas implies that their effect is potentially far reaching: even individuals who do not personally endorse beliefs that men are generally more competent than women are likely to be aware that these beliefs exist in the culture and expect that others will treat them according to these beliefs. This expectation, or what we think "most (other) people" believe, has been shown to modify behavior and bias judgments (Foschi 1996; Lovaglia et al. 1998; Steele 1997).

Double Standards for Assessing Competence

In an extension of status characteristics theory that is highly relevant for understanding how gender differences in self-assessments emerge, Foschi (1989) incorporates insights from the psychological literature on attribution to develop a theory about how individuals attribute performance to ability (or lack of ability). The main claim is that the standards individuals use to determine if a given performance is indicative of ability are a function of the diffuse status characteristics that are salient in a setting. When people who possess the lower state of a salient diffuse status characteristic (symbolized as D–) perform well at the group's task, their performances are critically scrutinized, because a good performance is inconsistent with status-based expectations for them. When people with the more valued state (D+) perform equally as well, their performances are consistent with expectations and are, therefore, less scrutinized. Since their performances are less scrutinized, higher status group members are judged by a more lenient standard than lower status group members. As a result, higher status group members are more likely to be judged as having task ability even when no "objective" performance differences exist. This result is predicted unless the task is one for which lower status individuals are believed to be "naturally" better, such as a task requiring nurturing ability in the case of gender. Empirical evidence supports these predictions for both gender (Foschi 1996) and race (Biernat and Kobrynowicz 1997).

When individuals assess their own competence at a task, they undoubtedly rely on performance information provided by legitimate evaluators (e.g., teachers, testing agencies, and employers). More positive evaluations of performance should lead to higher self-assessments of task competence. However, if we apply the double standard argument presented above in settings where individuals make assessments of their own competence, we would expect that, if gender is salient in the setting, gender will impact the performance expectations men and women hold for themselves. As long as the task is not one for which beliefs specifically advantage women, men will have higher performance expectations for themselves than otherwise similar women will. Men will, therefore, use a more lenient standard when assessing their own task competence. If individuals are provided with equal performance evaluations of their competence (e.g., have equal scores on a test), but men use a more lenient standard, then men will overestimate and women will underestimate their actual task ability. In this way, cultural beliefs about gender can lead to biased self-assessments of task competence. However, this prediction assumes that the status process just described occurs in the kind of setting where individuals assess their own competence.

Status Processes in Individual Settings

Applying status characteristics theory to develop this argument requires explaining why the theory should hold in settings where individuals commonly assess their own task competence. For instance, individuals likely assess their competence in settings where they take socially important mental ability tests, such as

intelligence quotient (IQ) tests, scholastic aptitude tests (SATs), and graduate record examinations (GREs). In these situations, which Erickson (1998) refers to as "individual evaluative settings," individuals are highly task oriented (i.e., they are focused on performing well on the task), but since they are not participating in a group, they are not collectively oriented. I first provide theoretical justification for why status generalization might occur in individual evaluative settings and then I review empirical evidence from the stereotype threat literature that is consistent with the theoretical argument.

THEORETICAL JUSTIFICATION. As previously mentioned, status characteristics theory has limited its scope to collectively oriented task groups, because in these groups the shared pressure to complete the group's task forces group members to anticipate the *relative* contribution of each member of the group. In other words, collective orientation produces the situational pressure to make relative comparisons about the future task performances of those in the setting. However, the logic of the theory does not specifically require collective orientation: instead it requires that some feature of the setting generate pressure for actors to consider their anticipated performance ability relative to others. Therefore, the theory should apply in other settings that generate this pressure, assuming that a status characteristic is salient, as described below.

Individual evaluative tasks represent another setting where individuals are potentially pressured to consider their performances relative to others. As Erickson (1998) describes, they will do so if they anticipate that they will receive a socially important and socially valid performance evaluation. This is because evaluative tasks, even if performed individually, often have the explicit purpose of ranking performances of actors. The use of evaluative tasks to rank individuals' performances is socially valid in the Weberian sense: individuals expect *others* to accept the ranking as legitimate and, consequently, orient their behavior towards this expectation (see Weber 1968:31–33). When individuals anticipate this ranking, they likely feel pressure to assess their task competence *relative* to others who, they imagine, are being or have been evaluated. This requires evaluating oneself in relation to the social environment. However, the standards for what constitutes a competent performance are not usually clearly defined beforehand, and others' precise scores are rarely known. In this uncertain environment, status characteristics, if salient, are available to influence performance expectations, as they do in collective task situations. Those with the more valued state (e.g., males, in the case of gender) should hold higher performance expectations and, therefore, see their performances as more competent than those occupying the less valued state (e.g., females), even if there are no differences in "objective" performances.

A status characteristic must be salient in the setting for this process to occur. Although individual settings, by definition, do not contain others to provide differentiation on a status characteristic, salience can be achieved if the characteristic is believed to be relevant to task performance (Wagner and Berger 1993). For example, if men (or women) are believed to be better at some task, gender will be a salient status characteristic. In addition to making gender salient, a belief about male superiority (or female superiority) at a particular task makes gender highly relevant to the goals of the setting, assuming individuals want to perform well on the task. Therefore, we would expect gender to impact individuals' behavior and assessments of their own competence in this setting. We would also expect the effect to be greater than if gender were perceived to be less relevant to the task at hand.

EMPIRICAL EVIDENCE. A growing body of empirical evidence is consistent with the idea that status generalization occurs in individual evaluative settings, such as those where individuals take socially important mentally ability tests. Lovaglia et al. (1998) demonstrate that individuals randomly assigned to low status conditions in experiments scored lower on a test of mental ability than those assigned to high status conditions. They contend that any attempt to measure mental ability needs to account for the way that salient status processes interfere with test-taking performance.

Similarly, the psychologist Claude Steele (1997) theorizes that individuals experience a self-evaluative threat in the presence of salient negative stereotypes about their group's intel-

lectual ability. This threat of social devaluation generates anxiety, arousal, and/or task-irrelevant processing that interferes with intellectual functioning and leads to decreased test performance (Steele and Aronson 1995). Although those who study "stereotype threat" use the more general term, "stereotype," these studies actually focus on the status element of stereotypes. They claim the belief that one category of the characteristic (African Americans, women) is less competent or capable than another (whites, men) causes the threat.[3]

Steele and Aronson (1995) show, for example, that when a difficult, standardized verbal exam is described as diagnostic of ability, African American students perform more poorly than white students. However, when the same test is not characterized as diagnostic of ability, African American and white students perform at the same level. Defining a test as ability-diagnostic primes a stereotype about race and verbal ability and makes race salient in the setting. Likewise, Shih, Pittinsky, and Ambady (1999) show that Asian women experience a stereotype threat, which reduces their mathematical performance when the stereotype that *women* have lower mathematical ability is primed, but they experience what might be called a stereotype "bonus," raising their mathematical performance, when the stereotype that *Asians* have superior mathematical ability is primed instead. As this study shows, the very same group of people, Asian women, can be advantaged or disadvantaged when performing the same task (a mathematics test) by varying what belief is described as relevant in the setting. Although beliefs about women and Asians are widely available in the culture, the impact of these beliefs in a given setting varies with the relevance of the belief in that setting, as status characteristics theory would predict.

Collectively, these studies indicate that status beliefs impact *task performance* in settings where individuals are task oriented, but are not members of a group. Based on these empirical results and the theoretical justification given, I make predictions about the effects of status beliefs on *self-assessments of task competence* in individual evaluative settings, identifying the conditions under which we would expect to see these effects.

Empirical Predictions

Based on the argument developed above, if status generalization occurs in individual evaluative settings, status beliefs will impact the self-assessments of individuals in these settings. This will occur when individuals are task oriented and anticipate that they will receive a socially important and socially valid performance evaluation. Under these conditions and if a diffuse status characteristic (D) is defined as relevant to the task at hand, performance expectations will vary positively with the state of D. Those with + states of D will have higher performance expectations than those with – states of D, assuming the task is not one for which cultural beliefs specifically advantage those with – states of D. In turn, higher performance expectations will lead to lower (more lenient) performance standards for inferring task ability (Foschi 1989). Therefore, I present the following hypotheses:

Hypothesis 1: In the presence of a belief making D task relevant and given an equal task outcome and differing states of D, those with a + state of D will assess their competence at the task higher than those with a – state of D.

Hypothesis 2: If competence at the task is perceived to be necessary for persisting on a particular career path, then higher self-assessments of competence lead to higher aspirations for activities that are associated with that career path.

When we apply hypothesis 1 to gender as a diffuse status characteristic, the hypothesis specifies the relationship between status beliefs about gender and the self-assessments men and women make of their task competence. When a status belief about a task advantages men (i.e.,

[3] One common criticism of the stereotype threat literature and the larger literature on stereotype activation is that there are inconsistencies in the mechanisms proposed to explain how stereotypes produce their effects depending on whether the stereotypes are negative or positive and whether the stereotypes are about one's own group or about other groups (see Wheeler and Petty 2001 for a review). Status characteristics theory, by contrast proposes one mechanism for the effects of status beliefs (regardless of whether the beliefs are advantaging or disadvantaging) on the behaviors and evaluations of both self and others.

if men are believed to be "naturally" better at a given task) and individuals are engaged in that task, gender is made salient and relevant to the goal of good performance. This primes a status generalization process where men make higher self-assessments of task competence compared to women who perform at the same level. Hypothesis 2 states the relationship between self-assessments and emerging aspirations for career-relevant activities. Although many factors influence individuals' preferences for various careers, I argue that cultural beliefs about gender differentially impact the emerging career-relevant aspirations of men and women.

THE EXPERIMENT

Overview

The experiment was designed primarily to evaluate hypothesis 1, regarding the bias that status beliefs impose on self-assessments of task competence. It will also provide some limited evidence about the effect of self-assessments on emerging aspirations. A test of the hypotheses calls for an experimental setting in which task performance can be controlled and the *relevance* of cultural beliefs about gender can be manipulated by associating or dissociating gender with task performance. The gender belief associated with the task and the gender of the subject are independent variables; self-assessment of task competence and the standard used to infer ability are the primary dependent variables. Task performance was experimentally held constant.

In one condition of the experiment, I manipulate gender belief associated with the task to advantage males (the "male advantaged" or "MA" condition). I provide subjects in this condition with evidence that males, on average, have more ability at the experimental task. This association between gender and task performance is intended to make gender salient and task relevant, leading to the prediction that men will use a more lenient standard than women when assessing their own task competence, resulting in higher male self-assessment levels. In the contrasting condition, I specifically dissociate gender beliefs from the task (the "gender dissociated" or "GD" condition) by providing subjects with evidence that there are no gender differences in task ability. The explicit dissociation of gender from the task should eliminate the task relevance of gender in the setting and, consequently, the effect of gender on self-assessments.

A more complete test of the bias imposed by gender status beliefs would also have included a condition where I present subjects with evidence that females have more task ability (i.e., a female advantaged condition). In this condition, I would predict women use a more lenient standard than men when assessing their own task ability, resulting in higher female self-assessments of task competence. Unfortunately, due to limited resources, only one experimental condition (the MA condition) and one contrasting condition (the GD condition) were run. However, the companion study described earlier provides some limited support for the female advantaged prediction. In that study, male students made significantly higher assessments of their mathematical ability than their equal-ability female counterparts. However, the effect was actually reversed when the students assessed their *verbal* ability: female students made significantly higher self-assessments of verbal ability, controlling for actual verbal performance (Correll 2001). To the extent that verbal tasks are culturally associated with women and students in the sample were aware of this cultural belief, the verbal result is consistent with the female advantaged prediction. Nonetheless, experimentally manipulating the gender association of the task to advantage women would have provided stronger evidence for evaluating the causal argument. A female advantaged condition would also have allowed us to test whether men and women respond differently to negative feedback about their own gender group. Also, it would have been useful in illuminating the source of the biasing effects that are found, as will be discussed in the results section below. In spite of these limitations, the two conditions presented here do provide the variation in the *relevance* of cultural beliefs about gender that is minimally necessary for evaluating the main hypothesis.

The final design, based on procedures drawn from Foschi (1996) and Erickson (1998), crosses the gender of subject with the "male advantaged" or "gender dissociated" presentation of the task, yielding four conditions. The subjects were male and female first-year undergraduates, who were paid for their time and randomly assigned to either the MA or GD condition. Analysis is based on a sample of 80 subjects (20

subjects per condition). A total of 94 subjects participated in the study, but I excluded the data from 14 of the subjects because they did not meet one of the scope conditions (n = 6)[4] or they were suspicious about some part of the study (n = 8).[5] Rejection rules were conservative and established beforehand. All analyses were also conducted with all available data and no substantive differences were found.

Procedures

Subjects came to the lab individually and were told that they were participating in the pretesting of a new set of graduate admissions examinations for a national testing service, purportedly measuring their "contrast sensitivity" ability. To establish that the task is socially important and would be used to legitimately rank "test-takers," as is required by the theory, subjects were informed verbally by an undergraduate experimenter and by reading a passage on their computer screen, that a national testing organization developed the contrast sensitivity exam and that both graduate schools and Fortune 500 companies have expressed interest in using this exam as a screening device. To further emphasize that individuals would be ranked based on their scores, subjects were also told that participants who scored in the top 25 percent of the scoring distribution would be entered into a drawing for a 50-dollar cash prize.

Next, the gender task belief manipulation was introduced. As a part of the initial verbal script delivered by the experimenter, participants were told either that males, on average, perform better on tests of contrast sensitivity (the MA condition) or that there is no gender difference in scores on tests of contrast sensitivity (the GD condition). To further emphasize the association or dissociation of gender with the task, subjects read more about gender and contrast sensitivity on their computer screen. In particular, the cover story described the interest of social science researchers in understanding either the gender difference or the lack of gender difference in performance on this task. This manipulation was intended to either make gender relevant to the goals of the situation (i.e., scoring high on the test) or to explicitly break the bond of relevance between gender and task performance.

Participants then completed two, 20-item rounds of the computer-administered contrast sensitivity test, in which subjects have five seconds to judge which color (black or white) predominates in each of a series of rectangles (Troyer 2001). The contrast sensitivity task is a reliable instrument commonly used in experimental social psychology. The task has no discernable right or wrong answers, yet subject suspicion in regard to the task is low (although see note 5). Since the amounts of white and black area are either exactly equal or very close to equal in each rectangle, it is impossible for subjects to actually derive correct solutions to the problems. All subjects were told that they correctly answered 13 of the 20 items during round one and 12 of 20 in round two. The scores were similar between rounds to convey that the test reliably measures contrast sensitivity ability. Mid-range scores, such as these, should allow for a wider range of self-assessment values than more extreme scores would (Foschi 1996). Giving all subjects identical test "scores" ensures that they assess their ability from objectively identical performance information.

After receiving their scores at the end of each round, participants answered a series of questions designed to first provide ability standard and then self-assessment measures. After the second round, they also answered a set of questions about how likely they would be to engage in activities that required high levels of task

[4] Six subjects answered "no" when asked if they wanted to do well on the task, thereby violating the scope condition that they be task-oriented. Of these six, 1 was a female from the GD condition, 1 was a female from the MA condition, 2 were males from the GD condition and 2 were males from the MA condition.

[5] The 8 cases excluded due to suspicion were 2 males and 2 females from the MA condition and 1 male and 3 females from the GD condition. The suspicion rate in this study is slightly higher than in comparable studies. This may be because the university where this study was conducted has students with some of the highest average SAT scores in the United States. As will be discussed in a subsequent section, all subjects were given the same, *average* score after they completed the experimental task. Students who were coded as being suspicious said that they did not think the task was "real" (i.e., they did not think the task had right and wrong answers) because they *personally* did not receive a high score on it. However, in analysis not shown, including their data in the analysis produced no substantive differences in the results.

ability. They then answered questions to assess the extent to which the experimental manipulations were successful. Before leaving, they were debriefed and paid.

Manipulation Checks

As is necessary for the experiment to provide an adequate test of the theory, items on the post-experiment computerized questionnaire evaluated whether the scope conditions of the theory were met and whether the manipulations of the gender belief associated with the task were successful. I phrased these questions in one of two ways. Some questions asked subjects how they thought "most people familiar with contrast sensitivity, including social science researchers, members of testing agencies and employers" view contrast sensitivity ("most people ratings"). Although the test was described as new, subjects learned that some groups of people were familiar with it and there was increasing interest in using the test or understanding this newly discovered ability. Other questions asked how subjects personally would describe contrast sensitivity ("personal ratings") after having been exposed to it. Subjects also completed a free response written questionnaire and were interviewed prior to debriefing as further checks on the experimental manipulations.

Scope Conditions. To evaluate the extent to which subjects believed that contrast sensitivity is a socially valued ability or that it is instrumental to other socially valued abilities (as the theory requires), I asked the subjects how "most people" would rate contrast sensitivity in terms of being important and predictive of success. For these and other questions, subjects used the mouse of their computer to drag a pointer between the two bipolar endpoints on a scale (e.g., "unnecessary" to "necessary"). The computer recorded a value between 0 and 100 to indicate how far to the right side of the scale the pointer was moved. Subjects indicated that most people would view the task as moderately important (mean 62.0, standard deviation [SD] 22.4) and predictive of success (mean 64.8, SD of 21.3).[6] Subjects also personally found the task to be moderately hard (mean 65.1, SD 12.4) and complex (mean 62.2, SD 17.7). Finally, subjects described the test as unfamiliar (mean 73.0, SD 18.5) and found the test instructions to be clear (mean 91.1, SD 10.8).

Gender Task Beliefs. To evaluate the manipulation of the relevance of gender in the setting, subjects rated how "most people" view contrast sensitivity on a 100-point scale ranging from "not at all masculine" to "highly masculine." Asking subjects about how they perceive that "most people" view the task, rather than how they personally understand it, is appropriate since status beliefs about gender are argued to impact behaviors and evaluations, even when individuals do not personally endorse the content of the belief (Ridgeway et al. 1998). However, for comparison, I asked subjects to provide their *personal* impressions of how masculine the task is.

Confirming the success of the manipulation, subjects in the MA condition indicated that "most people" view the task as significantly more masculine compared with their same gender counterparts in the GD condition.[7] As can be seen in Table 1, the mean "most people rating" for women in the MA condition was 62.1, compared with a rating of 24.4 for women in the GD condition ($t = 6.68$, $p < .001$). Likewise, the mean for males in the MA condition was 52.5, compared with a rating of 30.0 for their same gender counterparts in the GD condition ($t = 3.05$, $p < .01$). The differences between the means for males and females within condition are not significant. As with the "most people" ratings, both males and females in the MA con-

[6] In interviews conducted prior to debriefing, most subjects indicated that they viewed the task as *socially* important, meaning that they believed that the experimental "test," like other standardized tests, would have important consequences. Subjects often expressed this opinion with anger or frustration at what they described as the undue importance of standardized tests, in general, on individuals' life chances, but importantly they did see the experimental "test" in the category of those tests that "matter."

[7] In interviews before debriefing, subjects were asked questions designed to further assess whether the manipulation of the gender belief associated with the task was successful. Depending on condition, they were asked if they were surprised either that males had higher averages on the test (the MA condition) or that no gender difference exists for this test (the GD condition). Subjects in the MA condition vol-

Table 1. Means of Gender Task Association Manipulation Variables By Condition

	MA[a]		GD[a]	
	Females	Males	Females	Males
Most people ratings[b]				
Masculine	62.1 (16.8)	52.5 (24.9)	24.4 (18.8)	30.0 (21.6)
Feminine	39.4 (10.4)	36.0 (19.1)	31.4 (26.5)	40.0 (27.5)
Personal ratings[b]				
Masculine	51.4 (16.0)	45.7 (25.9)	21.4 (20.4)	26.4 (25.3)
Feminine	43.5 (12.1)	35.3 (20.9)	29.4 (26.7)	32.5 (28.2)

Note: Data shown as mean with standard deviation in parentheses; N = 80.

[a] Task beliefs: MA = male advantaged condition; GD = gender dissociated condition.

[b] Subjects were either asked how they perceived that "most people" would rate the task or how they "personally" would rate the task on a scale from "not at all masculine" (or feminine) to "highly masculine" (or feminine).

dition found the task to be significantly more masculine *personally* compared with their same gender counterparts in the GD condition.[8]

Because researchers have sometimes used a single continuum ranging from "masculine" to "feminine" to measure gender beliefs (c.f., Foschi 1996), I also asked subjects to rate how "most people" and they personally viewed the task on a scale ranging from not at all feminine to highly feminine.[9] I then performed a paired t-test within condition to compare the means of subjects' "most people ratings" on the masculine and feminine scales. Females and males in the MA condition had significantly higher "most people" masculine ratings than "most people" feminine ratings (females: $t = 7.62$, $p < .001$; males: $t = 2.80$, $p < .05$). In other words, they acknowledged that others were more likely to associate the task with masculinity than femininity. Females and males in the GD condition, by contrast, indicated that "most people" would associate the task more with femininity than masculinity (females: $t = -2.18$, $p = .024$; males: $t = 2.0$, $p = .058$). Together, these results offer considerable evidence that relevance of gender to task performance differed across conditions in the direction intended.

Measurement of Dependent Variables

Self-assessment. The main dependent variable is self-assessment of task ability, in this case contrast sensitivity ability. This variable is measured in three different ways. The primary self-assessment dependent variable is a composite variable constructed from subjects' responses to ten items on the computerized questionnaires in which they were asked to evaluate their performance on the contrast sensitivity test. Five contrasting adjective semantic differential items were included in the questionnaire following the first round and five identical items were included after the second round. The anchors for these items were: skilled/unskilled, incompetent/competent, incapable/capable, knowledgeable/unknowledgeable, adequate/inadequate. Subjects moved their computer mouse towards one of the two anchors to record values ranging from 0 to 100, a value of 100 indicating the extreme right anchor. The items were then recoded, if necessary, so that higher values represent higher assessments of ability.

unteered diverse hypotheses for the supposed gender difference, ranging from socialization to brain laterality explanations. In no instance, did a subject state that she or he did not believe that a gender difference exists. Before debriefing, subjects also filled out a free response questionnaire that asked them similar questions, which gave them an opportunity to express any suspicion privately. Although all subjects in the GD condition accepted the "finding" of no gender difference, some were surprised personally.

[8] The personal ratings are significantly lower than the "most people" masculine ratings for subjects in the MA condition (females: $t = 5.50$, $p < .001$; males: $t = 2.43$, $p < .05$); i.e., subjects view the task as less masculine "personally" compared to how they perceive "most (other) people" view it.

[9] Subjects had less differentiated ratings of how feminine they thought most people would view the task (Table 1). The differences between conditions are not significant. Not surprisingly, since the task beliefs were not manipulated to advantage females, all subjects found the task to be not very feminine.

The composite self-assessment variable is a factor score computed as a weighted -average generated by a confirmatory factor analysis in which all 10 items predict the latent variable "self-assessment." The fit statistics for this model indicate that the model is consistent with the data (GFI = . 914, CFI = . 980, RMSEA = . 084). The standardized factor loadings ranging from .60 to .96 suggest that the ten items measure the same underlying concept, in this case self-assessment of contrast sensitivity ability. The self-assessment composite has a mean of 47.7 and a SD of 13.0.

The second measure of self-assessment, the self-assessment "rating" variable, was constructed from an ordinal level question where participants were asked to describe their contrast sensitivity ability on a 7-point scale ranging from "considerably below average" to "considerably above average." This question was asked after both rounds of the study and responses were averaged. The mean of this item is 3.89 and its SD is 1.00.

The final self-assessment measure is taken from a single item in which participants were asked to assess how well they did on the contrast sensitivity test. This variable takes on a value between zero and 100 as participants moved their computer mouse between anchors of "not very well" to "very well." The mean of this item is 35.5 and the SD is 17.0. Not surprisingly, the three self-assessment variables (composite, rating, and single-item) are positively correlated with one another (Pearson correlations range from .52 to .67).

ABILITY STANDARD. After each round, participants indicated the score they would need to have achieved on the test to be convinced that they definitely possessed high levels of contrast sensitivity ability. Reponses from the two rounds were averaged.[10] The mean of the ability standard variable is 83.4 percent and the SD is 6.62 percent.

EMERGING ASPIRATIONS. To measure their emerging aspirations for activities described as career-relevant and requiring high task ability, I asked the subjects to rate (on a six-point scale ranging from "highly unlikely" to "highly likely") how likely they would be to: 1) apply to *graduate programs* requiring high levels of contrast sensitivity ability, 2) apply for a high-paying *job* requiring high levels of contrast sensitivity ability, 3) take a one-quarter *course* designed for those who possess high levels of contrast sensitivity ability in order learn more about this ability, and 4) enroll in a 3-hour *seminar* on contrast sensitivity for those with high contrast sensitivity ability.

Two composite variables were then created: one that is the sum of the course and seminar items and the other, the sum of the job and graduate school items, thereby allowing the composite variables to take on values from two to twelve. Confirmatory factor analysis provides support for this two-factor model of aspirations, with separate domains representing aspirations either *within* the university (courses and seminars) or *beyond* the university (graduate school and jobs). The mean of the "within university" aspiration variable is 7.49, with a SD of 2.57. The mean of the "beyond university" aspiration variable is 7.70, with a SD of 2.06.

These measures are intended to evaluate whether self-assessments impact *emerging aspirations* for activities that are believed to require task competence. Since participants only learned about the experimental task and ability upon arriving at the laboratory, we should not automatically assume that the measures reflect a commitment to actual behavior. It is also important to note that many factors will impact the aspirations individuals report for these or any activity. The argument is simply that the assessment individuals make of their own competence at a particular task will increase or decrease their emerging aspirations for paths requiring competence at that task. If men and women, on average, make different assessments of their own task competence, we would expect systematic gender differences in their aspirations for paths requiring some level of task ability.

[10] After the first round, subjects were asked to provide this score as a *percentage* of correct answers. After the second round, they were asked to provide the *number correct* out of twenty. Reponses from the second round were converted to a percentage scale and averaged with the first round value to create the ability standard variable.

RESULTS AND DISCUSSION

Status Beliefs and Biased Self-assessments Hypothesis

The main hypothesis about the biasing effect of gender status beliefs on self-assessments of task competence leads to the empirical prediction that men in the MA condition will assess their contrast sensitivity ability higher than women will. Recall that all subjects were given the same "score," thus ensuring that men and women received equal performance feedback. Men in the MA condition are also hypothesized to use a more lenient standard in assessing their task competence compared to women performing at the same level. No gender differences in self-assessments of task competence or in the ability standard used to assess competence are predicted in the GD condition, since the relevance of gender in the setting has been explicitly disassociated.

The top half of Table 2 provides means and SDs of the three self-assessment variables and the ability standard variable for women and men in the MA and GD conditions. In the MA condition, men have a mean self-assessment composite rating of 55.3 on a 100-point scale, and women, a rating of 41.1 of 100. The other two self-assessment variables show a similar pattern with men in the MA condition making higher self-assessments of their task competence than women in the MA condition. Men in the MA condition indicated that they would have to score at least 79.3 percent on a test of contrast sensitivity to be convinced that they had high task ability. Women reported that they would need a higher score of at least 88.9 percent correct to be certain they possessed high task ability. The gender differences in the means of these variables are smaller in the GD condition.

The results of a two-way (gender of subject and task belief) analysis of variance are shown in the lower half of Table 2. While no significant main effect is predicted for either the subject gender or the task belief factor, a significant interaction between the two factors is hypothesized since the interaction tests whether the effect of gender on self-assessments of task competence varies with the gender belief associated with the task. The interaction term, then, provides for a test of the status belief and biased self-assessment hypothesis. As can be seen in Table 2, the two-way interaction is significant for all four dependent variables, providing strong support for the main causal hypothesis.

Note that the gender main effect is also significant for three of the four dependent variables. This result is due to the magnitude of the experimental effect as can be seen in Table 3,

Table 2. Comparison of Means of Self-assessment and Ability Standard Variables by Subject Gender and Task Belief

	Self-assessment			Ability Standard
	Composite	Rating	Single item	
Means[a]				
MA Task Belief				
Female subjects	41.1 (13.5)	6.90 (1.59)	29.3 (15.2)	88.9 (4.61)
Male subjects	55.3 (14.8)	8.85 (2.43)	43.8 (20.0)	79.3 (7.70)
GD Task Belief				
Female subjects	47.1 (11.6)	7.60 (1.31)	35.5 (13.1)	82.4 (3.93)
Male subjects	47.2 (7.90)	7.75 (2.10)	33.4 (16.7)	83.1 (5.89)
F-Values[b]				
Factor				
Subject Gender	6.77*	6.05*	2.77	12.2**
Task Belief	0.146	0.220	0.331	1.09
2-way interaction	6.71*	4.45*	4.98*	16.6**

Note: N = 80.

[a] Data shown as mean with SD in parentheses; MA = male advantaged condition; GD = gender dissociated condition.

[b] F-values from the 2-way (gender × task belief) ANOVA.

* $p < .05$; ** $p < .01$

Table 3. Planned Contrast T-tests Comparing Means of Self-assessment and Ability Standard Variables Between Conditions

Dependent Variable	Contrast Number	Task Belief Contrast[a]	*t*
Self-assessment			
Composite	1	Female MA vs. Male MA	–3.17*
	2	Female GD vs. Male GD	–0.10
	3	Female MA vs. Female GD	–1.52
	4	Male MA vs. Male GD	2.17*
Rating	1	Female MA vs. Male MA	–2.73*
	2	Female GD vs. Male GD	–0.27
	3	Female MA vs. Female GD	–1.52
	4	Male MA vs. Male GD	1.18
Single item	1	Female MA vs. Male MA	–2.55*
	2	Female GD vs. Male GD	0.44
	3	Female MA vs. Female GD	–1.35
	4	Male MA vs. Male GD	1.78*
Ability standard	1	Female MA vs. Male MA	4.82**
	2	Female GD vs. Male GD	–0.47
	3	Female MA vs. Female GD	4.83**
	4	Male MA vs. Male GD	–1.78*

Note: N = 80.

[a] MA = male advantaged condition; GD = gender dissociated condition.

* $p < .05$; ** $p < .01$ (one-tailed, variances not assumed to be equal)

which contains the results of a series of planned contrast t-tests for pairs of conditions for each of the four dependent variables. Contrast 1 compares the means of the dependent variables for men and women in the MA condition and shows that men make significantly higher self-assessments of their task competence and use significantly lower ability standards than women when a task belief advantages men. In the gender-dissociated condition, however, the gender differences in self-assessments and ability standards are insignificant (see contrast number 2).

Consistent with the first hypothesis, when males are believed to be more competent at a task, men using a more lenient standard than women for assessing their own task competence, and consequently, men assess their own task ability higher than women performing at the same level. No significant gender difference was found in the GD condition, which suggests that men do not globally assess their task competence higher regardless of the task's gender association. Instead, when gender is made relevant in the setting, status beliefs about gender differentially bias the assessments men and women make.

SOURCE OF THE BIASING EFFECT. The results presented thus far demonstrate: 1) when a task belief advantages men a gender gap in self-assessment of task competence emerges that favors men, and 2) no gender gap appears when gender is specifically dissociated from the task. But, what is the source of the gender difference in the self-assessment in the male-advantaging condition? There are three logical ways that a gender gap in self-assessments can emerge. One is that men and women are both influenced, albeit in different directions and perhaps to differing degrees, by a male-advantaging task belief. In this scenario, confronted with a male-advantaging task belief men inflate their self-assessments and women deflate theirs, compared to the assessments they would have made if gender was explicitly defined as irrelevant to the task. The second possibility is that men ignore the male-advantaging task belief, but women attend to it. The final possibility is that women ignore the task belief, but men attend to it. To evaluate these possibilities, I compare mean self-assessments and ability standards for men and women in the MA condition with their same gender counterparts in the GD condition. These results are presented as the third and

fourth set of contrasts for each dependent variable in Table 3.

For both the main dependent variable—the composite self-assessment variable—and the single-item self-assessment variable, men in the MA condition made significantly higher assessments of their own task competence than did men in the GD condition. Furthermore, men in the MA condition were found to use a significantly more lenient standard for assessing their task ability than men in the GD condition (79.3 percent versus 83.1 percent). Taken together, these results suggest that men inflate their self-assessments of task competence when a task belief advantages them. This inflation of *self-assessments* in the presence of an advantaging task belief is similar to the stereotype bonus described earlier, where Asian women experienced enhanced mathematical *performance* when a positive stereotype about Asians and mathematical ability was primed (Shih et al. 1999).

The results for women were less consistent. Compared with women in the GD condition, women in the MA condition indicate that they would need higher scores to be certain they possessed high levels of task competence (i.e., their ability standard is higher). However, the differences between the means of the three self-assessment variables do not differ significantly between conditions for women. Taken together, the male and female results provide some evidence that both men and women are influenced by male-advantaging beliefs, although the results are more reliable for men. It is possible that men attend to male-advantaging beliefs more than women do. This result would be consistent with research showing that individuals are more prone to cognitive biases that are self-serving (Markus and Wurf 1987). If so, women should attend more than men to beliefs that are female-advantaging, a prediction that unfortunately could not be evaluated with the data collected in this study.

SUMMARY OF RESULTS. The main hypothesis was strongly supported. Men use a more lenient standard to infer ability and assess their task competence higher than women when exposed to a belief about male superiority, but no gender differences in self-assessments or ability standards were found when gender was defined as irrelevant to the task. Further, these differences were produced relatively easily. Although subjects had not heard of the task before participating in the study, after minimal exposure to a belief about male superiority and two rounds of testing, significant gender differences in self-assessments of task competence emerged. Finally, the results provide empirical support for the theoretical claim that status generalization occurs in individual evaluative settings under the conditions previously described.

The results from the experimental are consistent with those found in the analysis of the probability sample described earlier where male students assessed their own mathematical, but not verbal, competence higher than their equal ability female counterparts did. Mathematics is believed to be a masculine domain, while verbal skills are *not* culturally associated with masculinity (c.f., Hyde et al. 1990), thereby providing natural variation in the gender belief associated with these domains that is similar to that manipulated in the laboratory. Therefore, the survey results suggest that the causal mechanism evaluated with experimental data operates in a similar way in a "real world" (i.e., non-laboratory) setting.

The experimental data also rule out the alternative explanation described earlier for higher male self-assessments. Recall that, according to this alternative logic, self-assessments tap an additional component of unmeasured "real" ability, leading to the explanation that, in the case of mathematics, men make higher assessments of their mathematical ability, not because of the biasing effect of cultural beliefs about gender and mathematics, but because men "really are better" at mathematics. However, because correct solutions to the experimental task are impossible to derive, men cannot "really" be better at the experimental task. Nevertheless, when subjects, who were all given the same score on the task, were told that, on average, men perform better on the test, male subjects rated their task ability higher than female subjects did, consistent with the hypothesis advanced in this study.

In sum, both kinds of data support the hypothesis that status beliefs about gender bias the assessments men and women make of their own task competence. But, do gender differences in self-assessments influence emerging aspirations for activities that require task ability? In the next section, I evaluate the evidence designed to answer this question.

Emerging Aspirations Hypothesis

The second hypothesis is that higher self-assessments of competence at a particular task lead to higher aspirations for activities that require some level of competence at that task. Therefore, higher self-assessments of contrast sensitivity ability should be associated with higher levels of aspirations for future activities that are thought to require this ability, regardless of experimental condition. However, because men in the MA condition were found to assess their contrast sensitivity ability higher than women in this condition, these men should also have higher aspiration levels.

In Table 4, I provide a comparison of means and SDs of the two future aspiration variables by subject gender and task belief. The top portion of the table contains the means and SDs, and the lower portion provides results from two different 2-factor analysis of variance models. As expected, the means for the two future aspirations variables are higher for men in the MA condition than for women. However, women in the GD condition have higher mean aspirations than men. The two-way interaction between gender and task belief is significant for both aspiration variables in the 2-factor analysis of variance presented in the middle portion of Table 4. Consistent with theoretical predictions, the significant interaction term indicates the effect of gender on emerging aspirations does differ with the gender belief associated with the task. When a belief exists that men are better at a task, men have higher aspirations than women for paths requiring some level of task ability.

But, is the interactive effect the result of the gender difference in self-assessments found in the MA condition? To answer this question, I added the self-assessment composite variable as a covariate to the baseline model above (see bottom portion of Table 4). The self-assessment composite variable has a significant positive effect on both aspiration variables, and the model fit improves significantly with the addition of this variable. Higher self-assessments of task competence do increase individuals' report-

Table 4. Comparison of Means of Emerging Aspiration Variables by Subject Gender and Task Belief

	Within University Aspirations	Beyond University Aspirations
Means[a]		
MA Task Belief		
Female subjects	5.75 (1.89)	5.90 (1.62)
Male subjects	7.55 (3.19)	6.75 (2.47)
GD Task Belief		
Female subjects	6.75 (2.00)	7.65 (1.69)
Male subjects	5.90 (2.73)	6.70 (1.66)
F-Values[b]		
Factor		
Subject Gender	0.717	0.014
Task Belief	0.336	4.03*
2-way interaction	5.58*	4.52*
R-squared	.08	.10
F-Values[c]		
Factor		
Gender	0.040	0.620
Task Belief	0.251	4.69*
2-way interaction	2.83	2.05
Self-assess (beta)	0.052*	0.040*
R-squared	.14	.16

Note: N = 80.

[a] Data shown as mean with SD in parentheses; MA = male advantaged condition; GD = gender dissociated condition.

[b] F-values for baseline 2-factor ANOVA (gender × task belief), with no covariates.

[c] F-values for 2-factor ANCOVA (gender × task belief) with self-assessment covariate added.

* $p < .05$; ** $p < .01$

ed aspirations to continue on a path requiring high levels of task competence.

Importantly, the two-way interaction is no longer significant once the model is conditioned on level of self-assessment. This result is consistent with the mechanism advanced in this study: gender differences in self-assessments of task competence play a mediating role in producing gender differences in emerging aspirations. While many factors undoubtedly influence the formation of aspirations for activities thought to be career-relevant, the experimental data suggest that status beliefs about gender bias individual self-assessments and differentially influence the emerging aspirations of men and women.

Although the experimental data provide high-quality evidence for evaluating the biasing effect of gender status beliefs, the evidence in regard to emerging aspirations is more limited because the aspirations individuals express for activities associated with an unfamiliar task or ability are likely more tentative than they would be if the task or ability were more familiar. However, if the task or ability were more familiar, it would be more difficult to vary the gender beliefs associated with it, thereby making it harder to evaluate the main hypothesis about how gender beliefs bias self-assessments. (In fact, for some tasks, such as those requiring nurturing or mechanical ability, it would likely be impossible to convincingly vary the gender association of the task). Fortunately, while the companion study could only provide limited support for the main hypothesis, it is better suited for illustrating the effect of self-assessments on early career-relevant decisions.

Results from the companion study show that self-assessments of task competence do influence actual decisions that might be described as career-relevant. The fact that male students assessed their own mathematical competence higher than their equal ability female counterparts did explain part of the gender gap in enrollment in high school calculus courses and selection of a "quantitative" major. In this way, gender differences in self-assessment of mathematical competence, which are biased by cultural beliefs about gender and mathematics, influence actual commitment to paths leading to careers in science, math, and engineering, thereby contributing to the continued dearth of women in the quantitative professions. Taken together, the experimental and survey results illustrate that gender-differentiated self-assessments of task competence impact emerging aspirations and early career-relevant decisions.

SUMMARY

The main contribution of this study is to develop and evaluate a theoretical model that describes the constraining effect of cultural beliefs about gender on the emerging career-relevant aspirations of men and women. Using status characteristics theory and the empirical literature on stereotype threat, I argue that gender status beliefs will lead men and women to use different standards to judge their own task competence in individual evaluative settings, such as testing situations, when gender is salient and defined as relevant to performance in the setting. In this situation, I hypothesize that gender differences in self-assessments of task competence will emerge and lead to gender differences in emerging aspirations for career paths and activities that require task competence. The theoretical model is evaluated with data from an experiment that was designed to permit the manipulation of the relevance of gender in the setting, thereby providing for a strong test of the causal argument. Importantly, the experimental data support the model. A comparison of the experimental results with results from a probability sample illustrates the utility of the model and suggests that the causal process operates similarly in a "real word" setting.

More generally, the experimental results, along with the work on stereotype threat and the recent study by Lovaglia et al. (1998), indicate that the impact of status processes on the reproduction of inequality might be more far reaching than status characteristics theory has considered. Extending the scope of the theory to include individual evaluative settings, such as those described here, is an important advancement, since this setting is both very common and highly consequential in its impact on educational and occupational attainment. It includes most standardized test settings, including those that are used to determine college, graduate school, and professional school admissions and those used for certification in a wide range of professional occupations.

CONCLUSION AND IMPLICATIONS

The motivation for this study was to better understand how gender segregation in paid labor persists over other structural changes in society by focusing on the supply-side of the issue, examining how cultural beliefs about gender differentially constrain the emerging career-relevant aspirations or preferences of men and women. The implication of the theory is that if gender differences in aspirations emerge, men and women will likely make different career-relevant choices, which will funnel them into supply networks for different types of jobs. Rather than examining how men and women's aspirations emerge, many previous supply-side explanations simply document or assume that men and women *have* different aspirations or different career-relevant preferences. Economic models, in particular, tend to view aspirations or preferences as exogenous to labor market matching processes (England 1993). However, as I have shown, individuals form aspirations by drawing on perceptions of their own competence at career-relevant tasks, and the perceptions men and women form are differentially biased by cultural beliefs about gender. In this way, macro belief structures constrain emerging preferences and aspirations and, to the extent that individuals act on their aspirations, individual choice. The failure to recognize the constrained aspect of choice obscures some of the processes by which gender inequality is perpetuated. It either defines the problem away or locates its source in the individualistic actions of those already disadvantaged by their position in the labor market.

Shelley J. Correll *is Assistant Professor of Sociology and an affiliate with the Center for the Study of Inequality at Cornell University. She received her Ph.D. from Stanford University in 2001. Her research interests are in gender and social psychology, with the goal of explicating how various social psychological processes reproduce structures of gender inequality, especially in schools and the labor market. A current project explores how cultural understandings of the motherhood role create subtle discriminatory barriers for employed women who are mothers. Laboratory and field experiments are being conducted to evaluate the theoretical argument.*

REFERENCES

American Association of University Women. 1992. *How Schools Shortchange Girls*. Washington DC: American Association of University Women Educational Foundation.

Anker, Richard. 1997. "Theories of Occupational Segregation by Sex: An Overview." *International Labour Review* 136:315–39.

Berger, Joseph, Hamit Fisek, Robert Norman, and Morris Zelditch. 1977. *Status Characteristics and Social Interaction*. New York: Elsevier.

Biernat, Monica and Diane D. Kobrynowicz. 1997. "Gender and Race-based Standards of Competence: Lower Minimum Standards but Higher Ability Standards for Devalued Groups." *Journal of Personality and Social Psychology* 72: 544–57.

Bourdieu, Pierre. 1984 [1979]. *Distinction. A Social Critique of the Judgment of Taste*, translated by Richard Nice. Cambridge, MA: Harvard University Press.

Browne, Irene and Paula England. 1997. "Oppression from Within and Without in Sociological Theories: An Application to Gender." *Current Perspectives in Social Theory* 17: 77–104.

Conway, Michael, M. Teresa Pizzamiglio and Lauren Mount. 1996. "Status, Communality and Agency: Implications for Stereotypes of Gender and Other Groups." *Journal of Personality and Social Psychology* 71: 25–38.

Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-assessments." *American Journal of Sociology* 106: 1691–730.

Deaux, Kay and Mary Kite. 1987. "Thinking About Gender." Pp. 92–117 in *Analyzing Gender: A Handbook of Social Science Research*, edited by Beth Hess and Myra Marx Ferree. Newbury Park, CA: Sage.

Eccles, Jacquelynne S. 1994. "Understanding Women's Educational and Occupational Choices: Applying the Eccles et al. Model of Achievement-related Choices." *Psychology of Women Quarterly* 18: 585–609.

Eccles, Jacquelynne S., Bonnie Barber and Debra Jozefowicz. 1999. "Linking Gender to Educational, Occupational, and Recreational Choices: Applying the Eccles et al. Model of Achievement-related Choices." Pp. 153–192 in *Sexism and Stereotypes in Modern Society: The Gender Science of Janet Taylor Spence*, edited by William B. Swann, Jr., Judith H. Langlois, and Lucia Albino Gilbert. Washington D.C: American Psychological Association.

England, Paula. 1981. "Assessing Trends in Occupational Sex Segregation, 1900–1976." Pp. 273–95 in *Sociological Perspectives on the Labor Market*, edited by I. Berg. New York: Academic Press.

———. 1984. "Wage Appreciation and Depreciation: A Test of Neoclassical Economic Explanations of Occupational Sex Segregation." *Social Forces* 62:726–49.

———. 1992. *Comparable Worth: Theories and Evidence*. New York: Aldine.

———. 1993. "The Separative Self: Andocentric Bias in Neoclassical Economics." Pp. 37–53 in *Beyond Economic Man: Feminist Theory and Economics*, edited by Marianne A. Ferber and Julie A. Nelson. Chicago: The University of Chicago Press.

England, Paula, George Farkas, Barbara Stanek Kilbourne and Thomas Dou. 1988. "Explaining Occupational Sex Segregation and Wages: Findings from a Model with Fixed Effects." *American Sociological Review* 53:544–558.

Erickson, Kristan G. 1998. "The Impact of Cultural Status Beliefs on Individual Task Performance in Evaluative Settings: A New Direction in Expectation States Research." Ph.D. dissertation. Department of Sociology, Stanford University, Stanford, CA.

Fiske, Susan T. 1998. "Stereotyping, Prejudice, and Discrimination." Pp. 357–411 in *The Handbook of Social Psychology*, 4th edition (volume 2), edited by D.T Gilbert, S.T. Fiske, and G. Lindsey. Boston: McGraw-Hill.

Fiske, Susan T., Amy J.C. Cuddy, Peter Glick and Jun Xu. 2002. "A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow from Perceived Status and Competition. *Journal of Personality and Social Psychology* 82: 878–902.

Foschi, Martha. 1989. "Status Characteristics, Standards, and Attributions." Pp. 58–72 in *Sociological Theories in Progress: New Formulations*, edited by Joseph Berger, Morris Zelditch, Jr. and Bo Anderson.

———. 1996. "Double Standards in the Evaluation of Men and Women." *Social Psychology Quarterly* 59: 237–54.

Game, Ann and Rosemary Pringle 1983. *Gender at Work*. Boston: Allen and Unwin.

Glass, Jennifer. 1990. "The Impact of Occupational Segregation on Working Conditions." *Social Forces* 68:779–96.

Granovetter, Mark and Charles Tilly. 1988. "Inequality and Labor Processes." Pp. 175–220 in *The Handbook of Sociology*, edited by Neil J. Smelser. Newburry Park, CA: Sage Publications.

Hyde, Janet Shibley, Elizabeth Fennema, Marilyn Ryan, Laurie A. Frost and Carolyn Hoop. 1990. "Gender Comparisons of Mathematics Attitudes and Affect: A Meta Analysis." *Psychology of Women Quarterly*, pages 299–324.

Jacobs, Jerry A. 1989. "Long-term Trends in Occupational Segregation By Sex." *American Journal of Sociology* 95:160–73.

———. 1995a. "Trends in Occupational and Industrial Sex Segregation in 56 Countries, 1960–1980." Pp. 259–93 in *Gender Inequality at Work*, edited by Jerry A. Jacobs. Thousand Oaks, CA: Sage Publications.

———. 1995b. "Gender and Academic Specialties: Trends Among Recipients of College Degrees in the 1980s." *Sociology of Education* 68: 81–98.

Jacobsen, Joyce P. 1994. "Trends in Work Force Segregation, 1960–1990." *Social Science Quarterly* 75 (1): 204–11.

Lovaglia, Michael J., Jeffrey W. Lucas, Jeffrey A. Houser, Shane R. Thye, and Barry Markovsky. 1998. "Status Processes and Mental Ability Test Scores." *American Journal of Sociology* 104: 195–228.

Marini, Margaret Mooney and Mary C. Briton. 1984. "Sex Typing in Occupational Socialization." Pp. 192–232 in *Sex Segregation in the Workplace*, edited by Barbara Reskin. Washington DC: National Academy Press.

Markus, Hazel R. and Elissa Wurf. 1987. "The Dynamic Self-concept: A Social Psychological Perspective." *Annual Review of Psychology* 38: 299–337.

National Science Board. 1993. *Science and Engineering Indicators: 1993*. Washington DC: (NSB No. 87-1).

National Science Foundation. 1994. *Women, Minorities and Persons With Disabilities in Science and Engineering: 1994*. Arlington, VA (NSF 94-333HL).

Nelson, Robert L. and William P. Bridges. 1999. *Legalizing Inequality: Courts, Markets and Unequal Pay for Women in America*. Cambridge: Cambridge University Press.

Peterson, Trond and Laurie A. Morgan. 1995. "Separate and Unequal: Occupation-Establishment Sex Segregation and the Gender Wage Gap." *American Journal of Sociology* 101: 329–65.

Polachek, Solomon. 1976. "Occupational Segregation: An Alternative Hypothesis." *Journal of Contemporary Business* 5: 1–12.

———. 1981. "Occupational Self Selection: A Human Capital Approach to Sex Differences in Occupational Structure." *Review of Economics and Statistics* 58: 60–9.

Reskin, Barbara. 1993. "Sex Segregation in the Workplace." *Annual Review of Sociology* 19: 241–70.

Reskin, Barbara and Patricia A. Roos. 1990. *Job Queues, Gender Queues: Explaining Women's Inroads into Male Occupations*. Philadelphia: Temple University Press.

Ridgeway, Cecilia. 1997. "Interaction and the Conservation of Gender Inequality: Considering Employment." *American Sociological Review* 62: 218–35.

Ridgeway, Cecilia L., Kathy J. Kuipers, Elizabeth Heger Boyle and Dawn T. Robinson. 1998. "How Do Status Beliefs Develop? The Role of Resources and Interactional Experience." *American Sociological Review* 63: 331–50.

Ridgeway, Cecilia L. and Shelley J. Correll. 2000. "Limiting Gender Inequality Through Interaction:

The End(s) of Gender." *Contemporary Sociology* 29: (1) 110–20.

Shih, Margaret, Todd L. Pittinsky and Nalini Ambady. 1999. "Stereotype Susceptibility: Identity, Salience and Shifts in Quantitative Performance." *Psychological Science* 10: 80–3.

Steele, Claude M. 1997. "A Threat Is in the Air: How Stereotypes Shape Intellectual Identity and Performance." *American Psychologist* 52: 613–29.

Steele, Claude M. and J. Aronson. 1995. "Stereotype Threat and Intellectual Task Performance of African Americans." *Journal of Personality and Social Psychology* 69: 797–811.

Tienda, Marta and Vilma Ortiz. 1987. "Intraindustry Occupational Recomposition and Gender Inequality in Earnings." Pp. 23–51 in *Ingredients for Women's Employment Policy*, edited by Christine Bose and Glenna Spitze. Albany, NY: State University of New York Press.

Troyer, Lisa, 2001. *SES v 7.2: A Computerized Version of the Expectation States Research Program's Standardized Experimental Setting*.

Troyer, Lisa and Lesley C. Younts. 1997. "Whose Expectations Matter? The Relative Power of First-order and Second-order Expectations in Determining Social Influence." *American Journal of Sociology* 103: 692–732.

Wagner, David G. and Joseph Berger. 1993. "Status Characteristics Theory: The Growth of a Program." Pp. 23–63 and 454–63 in *Theoretical Research Programs: Studies in the Growth of Theory*, edited by Joseph Berger and Morris Zelditch Jr. Stanford, CA: Stanford University Press.

———. 1997. "Gender and Interpersonal Task Behaviors: Status Expectation Accounts." *Sociological Perspectives* 40: 1–32.

Weber, Max. 1968. *Economy and Society*, edited by G. Roth and C. Wittich and translated by E. Fischoff et al. New York: Bedminster Press.

Webster, Murray A., Jr. and Martha Foschi 1988. "Overview of Status Generalization." Pp. 1–20 in *Status Generalization: New Theory and Research*, edited by Murray A. Webster and Martha Foschi. Stanford, CA: Stanford University Press.

Wheeler, S. Christian and Richard E. Petty. 2001. "The Effects of Stereotype Activation on Behavior: A Review of Possible Mechanisms." *Psychological Bulletin* 127: 797–826.

Williams, John E. and Deborah L. Best. 1990. *Measuring Sex Stereotypes: A Multinational Study*. Newbury Park, CA: Sage.

Zellner, Harriet. 1975. "The Determinants of Occupational Pay." Pp. 44–70 in *Sex, Discrimination and the Division of Labor*, edited by Cynthia B. Lloyd. New York: Columbia University Press.